Appendix C. Source and Accuracy of Estimates

SOURCE OF DATA

The data were collected during the first eight interviews of the 1984 panel of the Survey of Income and Program Participation (SIPP). The SIPP universe is the noninstitutionalized resident population living in the United States. This population includes persons living in group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates and nursing home residents, were not eligible to be in the survey. Also, United States citizens residing abroad were not eligible to be in the survey. Foreign visitors who work or attend school in this country and their families were eligible; all others were not eligible. With the exceptions noted above, persons who were at least 15 years of age at the time of the interview were eligible to be interviewed in the survey.

The 1984 panel SIPP sample is located in 174 areas comprising 450 counties (including one partial county) and independent cities. Within these areas, clusters of two to four living quarters were systematically selected from lists of addresses prepared for the 1970 decennial census to form the bulk of the sample. To account for living quarters built within each of the sample areas after the 1970 census, a sample was drawn of permits issued for construction of residential living quarters through March 1983. In jurisdictions that do not issue building permits, small land areas were sampled and the living quarters within were listed by field personnel and then subsampled. In addition, sample living quarters were selected from a supplemental frame that included new construction for which building permits were issued prior to January 1, 1970, but for which construction was not completed until after April 1, 1970.

The first interview of this panel was conducted during October, November, and December 1983, and January 1984. Approximately one-fourth of the sample was interviewed in each of these months. Each sample person was visited every 4 months thereafter. At each interview, the reference period was the 4 months preceding the interview month.

Approximately 26,000 living quarters were originally designated for the sample. At the first interview, interviews were obtained from the occupants of about 19,900 of the 26,000 designated living quarters. Most of the remaining 6,100 living quarters were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. However, approximately 1,000 of the 6,100 living quarters were not inter viewed because the occupants refused to be interviewed, could not be found at home, were temporarily absent, or were otherwise unavailable. Thus, occupants of about 95 percent of all eligible living quarters participated in the first interview of the survey.

For subsequent interviews, only original sample persons (those inter viewed in the first interview) and persons living with them were eligible to be interviewed. Original sample persons were followed if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area. Then, telephone interviews were attempted. All first interview noninterviewed households were automatically designated as non interviews for all subsequent interviews. When original sample persons moved to remote parts of the country, moved without leaving a forwarding address or refused to be interviewed, additional noninterviews resulted.

A person was classified as interviewed or noninterviewed for the panel based on the following definitions. Interviewed sample persons were defined to be 1) those for whom self or proxy responses were obtained for each reference month of all eight interviews or 2) those for whom self or proxy responses were obtained for the first reference month of the panel and for each subsequent reference month until they were known to have died or moved to an ineligible address (foreign living quarters, institutions, or military barracks). Noninterviewed persons were defined to be those for whom neither self nor proxy responses were obtained for one or more reference months of the eight interviews (but not because they were deceased or moved to an ineligible address). All members of a household were excluded from the weighting procedure if one or more members had no self or proxy responses for the first interview. (The processing system was unable to handle persons in this type of household.)

Approximately 52,800 persons were counted as initially interviewed. This count excludes about 1,300 interviewed persons who were members of house holds in which one or more members were noninterviews in the first inter view. In the weighting procedure, approximately 32,400 persons were classified as interviewed. Persons who missed interviews due to the March 1985 sample cut were not classified as noninterviews but were adjusted for in the weighting procedure by a special factor (see "Estimation"). The person nonresponse rate is estimated to be 30 percent for the panel. Some respondents did not respond to some of the questions; therefore, the overall nonresponse rate for some items, especially sensitive income and money related items, is higher than the person nonresponse rate.

ESTIMATION

Several stages of weight adjustments were involved in the estimation procedure used to derive the SIPP longitudinal person weights. Each per son received a base weight equal to the inverse of his/her probability of selection. Two noninterview adjustment factors were applied. One adjusted the weights of interviewed persons in interviewed households to account for households which were eligible for the sample but could not be interviewed at the first interview. The second was applied to compensate for person noninterviews occurring in subsequent interviews. Another factor was applied to each interviewed person's weight to account for the SIPP sample areas not having the same population distribution as the strata from which they were selected.

An additional stage of adjustment to longitudinal person weights was per formed to reduce the mean square error of the survey estimates. This was accomplished by bringing the sample estimates into agreement with monthly Current Population Survey (CPS) type estimates of the civilian (and some military) noninstitutional population of the United States by age, sex, race, Hispanic ethnicity, and householder/not householder status as of the specified control date. The control date for the weights was November 1, 1983. The CPS estimates were themselves brought into agreement with estimates from the 1980 decennial census which have been adjusted to reflect births, deaths, immigration, emigration, and changes in the Armed Forces since 1980.

Spell estimates. Longitudinal person weights can be used to construct the average number of consecutive months of possession of a characteristic (i.e., the average spell length for a characteristic) during a given time period. For example, one could estimate the average length of each spell of receiving food stamps during 1985. Also, one could estimate the average spell of job search or layoff that elapsed before a person found a new job. To construct such an estimate, first identify the persons who possessed the characteristic at some point during the time period of interest. Then, create two sums of these person's appropriate longitudinal weights: (1) sum the product of the weight times the number of months the spell lasted and (2) sum the weights only. Now, the estimated average spell length is given by (1) divided by (2). A person who experienced two spells during the time period of interest would be treated as two per sons and appear twice in sums (1) and (2).

ACCURACY OF ESTIMATES

SIPP estimates are based on a sample; they may differ somewhat from the figures that would have been obtained if a complete census had been taken using the same questionnaire, instructions, and enumerators. There are two types of errors possible in an estimate based on a sample survey: nonsampling and sampling. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error. Found in the next sections are descriptions of sources of SIPP nonsampling error, followed by a discussion of sampling error, its estimation, and its use in data analysis.

Nonsampling variability. Nonsampling errors can be attributed to many sources, e.g., inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, inability or unwillingness on the part of the respondents to provide correct information, inability to recall information, errors made in collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, biases resulting from the differing recall periods caused by the inter viewing pattern used, and failure of all units in the universe to have some probability of being selected for the sample (undercoverage). Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers.

Undercoverage in SIPP results from missed living quarters and missed per sons within sample households. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed per sons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group. Further, the independent population controls used have not been adjusted for undercoverage. The Bureau has used complex techniques to adjust the weights for nonresponse, but the success of these techniques in avoiding bias is unknown.

Comparability with other estimates. Caution should be exercised when comparing data from this report with data from other SIPP publications or with data from other surveys. The comparability problems are caused by such sources as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures.

Sampling variability. Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence intervals. The sample estimate and its standard error enable one to construct confidence intervals, ranges that would include the average result of all possible samples with a known probability. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average result of all possible samples.
2. Approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average result of all possible samples.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

The average estimate derived from all possible samples is or is not contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence that the average estimate derived from all possible samples is included in the confidence interval.

Hypothesis testing. Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

All statements of comparison in the report have passed a hypothesis test at the 0.10 level of significance or better. This means that, for differences cited in the report, the estimated absolute difference between parameters is greater than 1.6 times the standard error of the difference.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be s_{DIFF}. If $X_A - X_B$ is between -1.6 times s_{DIFF} and +1.6 times s_{DIFF}, no conclusion about the characteristics is justified at the 10 percent significance level. If, however, $X_A - X_B$ is smaller than -1.6 times s_{DIFF} or larger than +1.6 times s_{DIFF}, the observed difference is significant at the 10 percent level. In this event, it is commonly accepted practice to say that the characteristics are different. Of course, sometimes this conclusion will be wrong. When the characteristics are, in fact, the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, if a large number of tests are performed, the significance of any single test should be inter preted cautiously.

Note concerning small estimates and small differences. Summary measures are shown in the report only when the base is 200,000 or greater. Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a base smaller than 200,000. Also, nonsampling error in one or more of the small number of cases providing the estimate can cause large relative error in that particular estimate. Estimated numbers are shown, however, even though the relative standard errors of these numbers are larger than those for the corresponding percentages. These smaller estimates are provided primarily to permit such combinations of the categories as serve each user's needs. Therefore, care must be taken in the interpretation of small differences since even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Standard error parameters and tables and their use. Most SIPP estimates have greater standard errors than those obtained through a simple random sample because clusters of living quarters are sampled for the SIPP. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required. Estimates with similar standard error behavior were grouped together and two parameters (denoted "a" and "b") were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These "a" and "b" parameters vary by characteristic and by demographic subgroup to which the estimate applies. Table C-1 provides base "a" and "b" parameters to be used for estimates in this report.

For those users who wish further simplification, we have also provided general standard errors in tables C-2 and C-3. Note that these standard errors must be adjusted by a factor from table C-1. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors are given in the following sections.

It should be noted that the parameters given in table C-1 for calculating standard errors of spell estimates are preliminary and should be used cautiously. In general, the use of the spell estimate parameters may result in underestimates of standard errors. Research is currently underway to further evaluate the accuracy of parameters used for calculating standard errors of spell estimates presented in this report.

Standard errors of estimated numbers. The approximate standard error, s_x, of an estimated number of persons shown in this report can be obtained in two ways.

It may be obtained by the use of the formula

$$s_x = fs \qquad (1)$$

where f is the appropriate "f" factor from table 1, and s is the standard error of the estimate obtained by interpolation from table 2. Alternatively, s_x may be approximated by the formula

$$s_x = \sqrt{ax^2 + bx} \qquad (2)$$

Here x is the estimated number and "a" and "b" are the parameters associated with the particular type of characteristic. Use of formula (2) will provide more accurate results than the use of formula (1).

Illustration. Suppose the SIPP estimate of the total number of persons experiencing only one spell of job search or layoff that began in 1984 is 17,296,000. The appropriate "a" and "b" parameters to use in calculating a standard error for the estimate are obtained from table C-1. They are a= -0.0000322 and b = 5,800, respectively. Using formula (2), the approximate standard error is

$$\sqrt{(-0.0000322)(17{,}296{,}000)^2 + (5{,}800)(17{,}296{,}000)} = 301{,}000$$

Table C-1. **SIPP Generalized Variance Parameters for Estimates**

Subject	a	b	f factor
PERSONS WITH SPELLS			
Total or White:			
Both Sexes	-0.0000322	5,800	1.00
Male	-0.0000672	5,800	1.00
Female	-0.0000617	5,800	1.00
Black:			
Both Sexes	-0.0002809	7,804	1.16
Male	-0.0005994	7,804	1.16
Female	-0.0005283	7,804	1.16
NUMBER AND LENGTH OF SPELLS			
Total or White persons:			
Both Sexes	-0.0000464	8,353	1.20
Male	-0.0000967	8,353	1.20
Female	-0.0000888	8,353	1.20
Black persons:			
Both Sexes	-0.0004044	11,238	1.39
Male	-0.0008632	11,238	1.39
Female	-0.0007608	11,238	1.39

The 90-percent confidence interval as shown by the data is from 16,814,000 to 17,778,000. Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all samples.

Using formula (1), the appropriate "f" factor (f=1.00) from table C-1, and the standard error of the estimate by interpolation using table C-2, the approximate standard error is

$$s_x = (1.00)(301{,}000) = 301{,}000$$

The 90-percent confidence interval as shown by the data is from 16,814,000 to 17,778,000.

Standard errors of estimated percentages. This section refers to the two types of percentages represented in this report. These are the percentages of a group of persons possessing a particular attribute and the percentages of spells of job search and layoff. For example, the percentage of persons who began spells of job search and layoff in 1984 and the percentage of spells of job search and layoff that began and ended with a job in 1984 demonstrates the two types of percentages, respectively. The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percent age is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are over 50 percent. For example, the percent of employed persons is more reliable than the estimated number of employed persons. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

For the percentage of persons or spells, the approximate standard error, $s_{(x,p)}$, of the estimated percentage, p, can be obtained by the formula

$$s_{(x,p)} = fs \tag{3}$$

where f is the appropriate "f" factor from table C-1, and s is the standard error of the estimate obtained by interpolation from table C-3. Alternatively, it may be approximated by the formula

$$s_{(x,p)} = \sqrt{\frac{b}{x}p(100\text{-}p)} \tag{4}$$

Here x is the base of the percentage, p is the percentage ($0<p<100$), and b is the "b" parameter associated with the characteristic in the numerator. Use of this formula will give more accurate results than use of formula (3).

Illustration. Suppose that we have a SIPP estimate of 5,837,000 spells of job search for males aged 25 to 54. Of these, 85.0 percent ended in a job. Using formula (4) and the "b" parameter of 8,353 (from table C-1), the approximate standard error is

$$\sqrt{\frac{(8{,}353)}{(5{,}837{,}000)}(85.0)(100\text{-}85.0)} = 1.4 \text{ percent}$$

Consequently, the 90-percent confidence interval as shown by these data is from 82.8 to 87.2 percent.

Using formula (3), the appropriate "f" factor ($f=1.20$) from table C-1, and the appropriate s by interpolation using table 3, the approximate standard error is

$$s_x = (1.2)(1.09) = 1.3 \text{ percent}$$

Table C-2. **Standard Errors of Estimated Numbers**

(Numbers in thousands)

Size of estimate	Standard error	Size of estimate	Standard error
200	34	22,000	335
300	42	26,000	359
600	59	30,000	387
1,000	76	50,000	458
2,000	107	80,000	508
5,000	169	100,000	508
8,000	211	130,000	458
11,000	245	135,000	443
13,000	264	150,000	381
15,000	282	160,000	321
17,000	299	180,000	27

The 90-percent confidence interval shown by these data is from 82.9 to 87.1 percent.

Standard error of a mean or aggregate. A mean is defined here to be the average quantity of some characteristic (other than the number of per sons) per person. An aggregate is defined to be the total quantity of some characteristic summed over all units in a subpopulation. For example, a mean could be the average number of spells of females age 25 to 54; an aggregate, the total spells for that subpopulation. The standard error of a mean can be approximated by formula (5) below and the standard error of an aggregate can be approximated by formula (6). Because of the approximations used in developing formulas (5) and (6), an estimate of the standard error of the mean or aggregate obtained from these formulas will generally underestimate the true standard error. The formula used to estimate the standard error of a mean, x, is

$$s_x = \sqrt{\frac{b}{y}s^2} \tag{5}$$

where y is the base, s^2 is the estimated population variance of the characteristic and b is the "b" parameter associated with the particular type of characteristic. The standard error of an aggregate k is estimated by:

$$s_k = \sqrt{b\, y\, s^2} \tag{6}$$

To estimate the population variance, s^2, the range of values for the characteristic is divided into c intervals, where the lower and upper boundaries of interval j are Z_{j-1} and Z_j, respectively. Each person is placed into one of the c groups such that the value of the characteristic is between Z_{j-1} and Z_j. The estimated population variance, s^2, is then given by formula (7):

$$s^2 = \left[\sum_{j=1}^{c} p_j m_j^{\,2}\right] - \bar{x}^2 \tag{7}$$

where p_j is the estimated proportion of persons in group j (based on weighted data), and $m_j = (Z_{j-1} + Z_j)/2$. The most representative value of the characteristic in group j is assumed to be m_j. If group c is open ended, i.e., no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2}Z_{c-1}$$

The mean, $\bar{x}$, can be obtained using the following formula:

$$\bar{x} = \sum_{j=1}^{c} p_j m_j \tag{8}$$

Illustration for mean. Suppose that the distribution of the number of spells are given in the following table for women aged 25 to 54 who were looking for jobs or on layoff that began in 1984 and were completed.

Subject	Total	1 spell	2 spells	3 spells	4 spells or more
Females, 25-64 years (thous.)	5,988	4,871	960	129	29
Percent of total	100	81.4	16.0	2.2	0.4

The average number of spells per person from formula (8) is

$$\bar{x} = 0.814\ (1) + 0.16\ (2) + 0.022\ (3) + 0.004\ (6) = 1.22$$

Using formula (7) and the average number of spells per person of 1.22 the estimated population variance, s^2, is

$$s^2 = [30.814(1)^2 + 0.16(2)^2 + 0.022\ (3)^2 + 0.004\ (6)^2] - (1.22)^2 = 0.3076$$

The appropriate "b" parameter from table 1 is 5,800. Now using formula (5), the estimated standard error of the mean is

$$s_x = \sqrt{\frac{5{,}800}{5{,}988{,}000}(0.3076} = 0.0176$$

The 90-percent confidence interval for the mean shown by these data is from 1.19 to 1.25 spells of job search and layoff per women aged 25 to 54 years.

Illustration for aggregates. Suppose the total number of spells for this subpopulation is 7,293,000. Now, using formula (6) and the appropriate "b" parameter 5,800 from table 1, the estimated standard error of the aggregate is

$$s_k = \sqrt{(5{,}800)(5{,}988{,}000)(0.3076)} = 103{,}000$$

The 90-percent confidence interval for the aggregate shown by these data is 7,128,000 to 7,458,000 total number of spells for women aged 25 to 54 years.

Standard error of a difference. The standard error of a difference between two sample estimates, x and y, is equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2 - 2rs_xs_y} \qquad (9)$$

where s_x and s_y are the standard errors of the estimates x and y, and *r is the correlation coefficient between the characteristics estimated by x and y. The estimates can be numbers, averages, percents, ratios, etc. When the characteristics estimated by x and y were considered correlated in this report, a correlation coefficient of 0.5 was used in the above formula. This value is considered preliminary and should be used cautiously. Underestimates or overestimates of standard error of differences result if the estimated correlation coefficient is overestimated or underestimated, respectively.

Illustration. Suppose that we are interested in the difference in the number of men and women who received health insurance coverage during their first and only spell of job search. Of the men and women aged 25-54 who were covered by health insurance one month before their spell began, 1,743,000 men remained covered through all months of their spell compared to 2,526,000 women who were covered. Using parameters and factors from table C-1, the standard errors of these numbers are approximately 100,000 and 119,000 for men and women, respectively.

Table C-3. **Standard Errors of Estimated Percentages**

Base of estimated percentage (thousands)	Estimated percentage					
	≤1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200	1.7	2.4	3.7	5.0	7.2	8.5
300	1.4	2.0	3.0	4.0	5.9	7.0
600	1.0	1.4	2.1	2.9	4.2	4.9
1,000	0.8	1.0	1.7	2.3	3.2	3.8
2,000	0.5	0.8	1.1	1.6	2.3	2.7
5,000	0.3	0.4	0.7	1.0	1.5	1.7
8,000	0.3	0.4	0.6	0.8	1.1	1.4
11,000	0.2	0.3	0.5	0.7	1.0	1.2
13,000	0.2	0.3	0.5	0.6	0.9	1.1
17,000	0.2	0.3	0.4	0.5	0.8	0.9
22,000	0.2	0.2	0.4	0.5	0.7	0.8
26,000	0.2	0.2	0.3	0.5	0.7	0.8
30,000	0.1	0.2	0.3	0.4	0.6	0.7
50,000	0.1	0.2	0.2	0.4	0.5	0.5
80,000	0.1	0.1	0.2	0.3	0.4	0.4
100,000	0.1	0.1	0.2	0.2	0.3	0.4
130,000	0.1	0.1	0.2	0.2	0.3	0.3
180,000	0.1	0.1	0.1	0.2	0.3	0.3

Now, the standard error of the difference is computed using the above two standard errors. The correlation between these estimates is zero. Therefore, standard error of the difference is computed by formula (9):

$$s_{(x-y)} = \sqrt{(100{,}000)^2 + (119{,}000)^2} = 155{,}000$$

Suppose that it is desired to test at the 10 percent significance level whether the number of men differs significantly from the number of women. To perform the test, compare the difference of 783,000 to the product 1.6 x 155,000 = 248,000. Since the difference is larger than 1.6 times the standard error of the difference, the data show that the estimates for the number of men and women covered by health insurance one month before and for every month during their spell differ significantly at the 10 percent level.